国家标准《统一内容标签格式规范》介绍

（杨鹏，副教授，东南大学未来网络研究中心副主任）

2017 年 12 月 29 日，由全国中文新闻信息标准化技术委员会牵头制定的中华人民共和国国家标准《统一内容标签格式规范》，经国家质量监督检验检疫总局、国家标准化管理委员会正式批准颁布，标准编号 GB/T 35304–2017。该标准的主要起草单位包括东南大学、新华通讯社、国家新闻出版广电总局、清华大学、大有数字资源有限公司和北京大学。统一内容标签（Uniform Content Label，简称 UCL）是我国科学家原创的内容元数据，国家标准《统一内容标签格式规范》的正式颁布，为统一内容标签技术的推广、普及和应用，以及内容大数据的规范标引、高效共享与依法治理等，提供了关键的标准支持。

一、统一内容标签国标研制背景

在 2000 年左右，中国工程院李幼平院士在深入研究互联网信息共享方式及其缺陷的基础上，提出了统一内容标签 UCL 的概念。关于 UCL 的最初研究成果主要发表在《中国工程科学》和《电视技术》等杂志上。2002 年，《中国传媒科技》陆续发表了李幼平院士和新华通讯社孙宝传同志关于后万维网时代和第五媒体“猜想”的对话，对 UCL 及新型信息共享网络进行了早期较为系统的阐述。自统一内容标签 UCL 概念提出以来，经过李幼平院士与多位科学家、领域专家和研究人员等的多年研究和应用实践，它的内涵和外延得到不断丰富和完善。

统一内容标签主要在统一资源定位符 URL（Uniform Resource Locator）的基础上发展而来。URL 是互联网的基础，互联网中的各种信息资源都是按照 URL 地址符（“Locator”）进行组织。URL 的提出及标准化极大推动了互联网的发展，但是 URL 通常只能表示信息资源的位置，无法描述信息资源的丰富语义，由此带来互联网信息资源难找、难管、失序等弊端。具体而言，互联网的主要弊端表现为：有用信息不易寻找，个性需求无法满足；网上内容难以有效治理，良莠不齐垃圾泛滥；网络导读严重缺失，舆论导向难以落实等。尽管 Tim Berners–Lee 提出了语义网（Semantic Web）概念，试图使 Web 变成能够自动理解词语和概念、以及它们之间逻辑关系的智能网络，但是语义网实现起来非常困难。统一内容标签从互联网中内容资源难找、难管和失序等问题的根本症结入手，兼顾了内容共享应用中的三个重要角色（读者、作者和管理者），能够有效弥补统一资源定位符 URL 的语义缺失和管理缺失。

关于统一内容标签的研究，与我国近年来若干重大战略需求紧密相关。互联网已经成为网络空间的主要载体，与此同时，互联网的缺陷和不足日益显露。我国政府发出“用好互联网，管好互联网”的号召，提倡采用“三网融合”营造国家公共文化服务体系。2013 年 11 月，全面深化改革若干重大问题的决定，指出要“促进基本公共文化服务标准化、均等化”，在此基础上“构建现代公共文化服务体系”；同时，要“确保国家网络和信息安全”。2014 年 2 月，中央网络安全和信息化领导小组第一次会议指出，没有网络安全就没有国家安全，没有信息化就没有现代化，网络安全和信息化是一体之两翼、驱动之双轮，我们要从网络大国建设成为网络强国。2015 年 12 月，第二届世界互联网大会指出，依法开展网络空间治理，网络空间日渐清朗，国际社会应该加强对话合作，推动互联网全球治理体系变革，构建网络空间命运共同体。2016 年 4 月，网络安全和信息化工作座谈会指出，网络空间是亿万民众共同的精神家园，网络空间天朗气清、生态良好，符合人民利益，网络空间不是“法外之地”，要依法加强网络空间治理。统一内容标签充分考虑这些战略需求，采取内容驱动的先进理念对内容标识进行了全新设计，形成生产、消费和管理三位一体的内容大数据创新标识体系，使之成为网络空间新的内容基元。

此外，统一内容标签的研究也与媒体领域的最新发展趋势紧密关联。2015 年 6 月，苹果公司在 Apple 全球开发者大会上推出 Apple News，它可以为 iOS 用户提供新闻内容定制服务，热门网站和报纸的内容都可以按用户需求出现在 Apple News 中；Facebook 也在同时期推出了交互式媒体内容创建和发布工具 Instant Articles。2015 年 10 月，国际著名媒体纽约时报（New York Times）为了应对苹果公司和 Facebook 等进军传媒业的挑战，提出研制“新闻编码”（Particles Code），通过可检索的“新闻编码”来实现累积性新闻（accumulative news），把新闻报道按照时间轴和知识点进行组织。我国政府高度重视互联网时代媒体融合与转型的新趋势，中央《关于推动传统媒体和新兴媒体融合发展的指导意见》明确提出，坚持传统媒体和新兴媒体优势互补、一体发展，推动传统媒体和新兴媒体在内容、渠道、平台、经营、管理等方面的深度融合。统一内容标签顺应传统媒体、新兴媒体、以及互联网等领域的发展趋势，按照媒体融合发展要求着力支持内容信息的聚合、传播、管理和个性化服务等，注重提高海量内容的汇聚和分析能力，建立“内容自寻网民”的主动服务机制。

东南大学李幼平院士团队关于统一内容标签的研究工作，受到全国中文新闻信息标准化技术委员会的高度重视，从 2013 年开始推动统一内容标签的标准化工作。在此过程中，国家若干重要战略需求和媒体领域最新发展趋势得到充分重视，并通过不断发展和完善，体现在刚刚颁布的统一内容标签国家标准中。

二、统一内容标签国标研制过程

考虑到统一内容标签自 2000 年以来已经过十余年研究和发展，其核心思想和主体结构已趋成熟，李幼平院士从 2013 年初开始筹划统一内容标签的标准化工作，并带领东南大学未来网络研究中心团队进行了大量准备工作。2013 年 11 月，在全国中文新闻信息标准化技术委员会第二次主任办公会上，明确提出“筹建‘统一内容标签标准工作组’，加快统一内容标签技术标准的制定”的要求。2014 年 2 月，新标委秘书处邀请李幼平院士担任工作组顾问，正式成立“统一内容标签标准工作组”，致力于统一内容标签技术标准的制定和推广工作。工作组主要由东南大学、新华通讯社、国家新闻出版广电总局、清华大学、大有数字资源有限公司和北京大学等机构的专家学者组成。

统一内容标签标准工作组自成立后，迅速开始着手研制统一内容标签的核心标准《统一内容标签（UCL）格式规范》。2014 年 3 月，统一内容标签标准工作组召开第一次会议，讨论和审议了工作组关于统一内容标签（UCL）研究进展及标准研制工作的报告，围绕统一内容标签中代码部分和属性部分的格式展开讨论，并达成一致。在此基础上，工作组整理汇总了前期 UCL 技术的主要研究成果，组织召开了多次讨论会，并对国内外相关技术标准进行深入研究，在充分吸收各方意见建议后，由多名起草专家协同工作、联合攻关，于 2014 年 5 月初完成《统一内容标签（UCL）格式规范（征求意见稿）》的起草工作。随后又进行了广泛的意见征求，经修改完善后形成标准送审稿，于 2014 年 8 月顺利通过标准审查会的审查，并于 2015 年 1 月正式发布成为新华社技术标准。

以新华社技术标准《统一内容标签（UCL）格式规范》为基础，2015 年 3 月新标委秘书处上报国家标准化管理委员会申请国家标准立项，国家标准化管理委员会在 2015 年 7 月正式批准 UCL 国家标准立项。其后的两年多时间，工作组围绕统一内容标签国家标准编制工作，多次召开工作组会议和院士专家咨询研讨会，不断吸纳媒体融合发展、网络空间安全、互联网内容治理等新的需求，从 UCL 描述完备性和结构灵活性方面对标准文本进行了创新性修改和完善，先后完成了统一内容标签推荐性国家标准复审、国家标准《统一内容标签格式规范》征求意见稿、送审稿和报批稿，在 2017 年 5 月报送国家标准化管理委员。2017 年 12 月 29 日，经国家质量监督检验检疫总局、国家标准化管理委员会批准，《统一内容标签格式规范》正式颁布成为中华人民共和国国家标准，标准编号为 GB/T 35304–2017，标准正式实施时间为 2018 年 4 月 1 日。

三、统一内容标签国标主要内容和特色

国家标准《统一内容标签格式规范》按照科学、完善的方法定义了 UCL 数据包的基本格式，包括各个域的名称、含义及定义规则等。UCL 数据包主要分成前后两个部分，分别为 UCL 代码（UCL Code）部分和 UCL 属性（UCL Properties）部分。UCL 代码部分储存与内容相关的代码信息，用于面向读者的快速导航，主要的域包括版本、媒体类型、优先级与策略、内容来源、一级类别、二级类别、话题、内容版权与类型、安全能级码、语种、时间戳、顺序号等，标准规约了各个域的定义、位置、长度、取值、数据类型等格式信息。UCL 属性部分储存与内容相关的语义信息和管理信息等，主要的域包括 UCL 属性总头部、内容标题、内容关键词、内容摘要、内容作者、内容实体、传播路径、内容数字签名、内容责任链、UCL 包数字签名等，标准规约了各个域的定义、属性类别、属性长度、属性净荷、净荷数据类型等格式信息。

国家标准《统一内容标签格式规范》瞄准我国未来信息共享网络的发展目标，聚焦互联网中内容难找、难管和失序等棘手问题，以互联网思维和大数据思维为指导，为整合富类型、多渠道、宽领域的国家信息资源提供标准支持，成为富有中国特色的原创性内容大数据创新标识体系。大数据趋势与互联网飞速发展下的媒体变革，已经促使互联网演变为“博大精深”内容大数据的集散地，实现全民科学享用内容大数据，必然面临三个挑战：内容共享如何普惠全民，内容服务如何安全可信，国家如何对内容依法治理。统一内容标签国家标准是应对这些挑战的利器，它可以提高内容大数据的聚合和分析能力，提升内容的传播力和影响力，建立“以内容为中心”的普惠全民引导机制，为实现“加快构建舆论引导新格局”做出贡献。

国家标准《统一内容标签格式规范》聚焦和服务于我国现代公共文化服务体系、网络空间安全与媒体融合发展等重大战略需求，填补了我国传媒科技领域和原创性内容元数据标准方面的重大空白。统一内容标签国家标准适用于对传媒、互联网及相关行业中主要面向群体受众的内容资源进行标引，其应用范围不仅包括新闻网站、信息门户、通讯社、广播、电视、报刊、网络媒体等新闻媒体，也包括应急广播、内容聚合、舆情分析、版权保护等其他应用，并且已经在相关领域得到实际应用。与现有的各种内容元数据标准和内容标识技术方案相比，统一内容标签具有理念先进、简洁高效、结构灵活等特点，可以从多个维度描述内容资源的语用、语义和管理信息，允许按照应用需求进行裁剪和自由扩展，能够有效支持内容大数据的高效聚合与泛在分发、个性化主动服务、语义分析与知识萃取、认证注册物证链管理、追踪溯源与依法追责等，尤其在内容依法治理与网络空间安全等方面具有显著优势。